

Variaciones en el proceso de confirmación diagnóstica entre unidades de cribado poblacional de cáncer de mama



Carmen Natal^{a,*}, Ana Fernández-Somoano^{b,c,d}, Isabel Torá-Rocamora^{d,e}, Adonina Tardón^{b,c} y Xavier Castells^{d,e}, en nombre del grupo BELE[◇]

^a Servicio de Salud del Principado de Asturias, Oviedo (Asturias), España

^b CIBER de Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Madrid, España

^c Universidad de Oviedo, Oviedo (Asturias), España

^d IUOPA-Área de Medicina Preventiva y Salud Pública, Departamento de Medicina, Universidad de Oviedo, Asturias, España

^e Departamento de Epidemiología y Evaluación, IMIM (Hospital del Mar Instituto de Investigación Médica) y Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Barcelona, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Recibido el 12 de enero de 2016

Aceptado el 15 de marzo de 2016

On-line el 29 de abril de 2016

Palabras clave:

Cáncer de mama

Detección precoz de cáncer

Técnicas y procedimientos diagnósticos

Variaciones en la práctica médica

Análisis de áreas pequeñas

Evaluación de servicios sanitarios

R E S U M E N

Objetivo: Analizar las variaciones en el proceso de confirmación diagnóstica entre unidades de cribado, las variaciones en los resultados de cada episodio y la relación entre las tasas de las diferentes pruebas de confirmación diagnóstica y las de detección de lesiones.

Método: Estudio observacional de las variaciones de las tasas estandarizadas de pruebas diagnósticas y de detección de lesiones, en 34 unidades de cribado de los programas poblacionales de detección precoz de cáncer de mama, de tres comunidades autónomas, en el periodo de 2002 a 2011.

Resultados: Las razones de variación entre los percentiles 25-75 en las tasas de realización de pruebas diagnósticas oscilaron entre 1,68 (recitaciones) y 3,39 (punción-aspiración con aguja fina). En las tasas de detección de lesiones benignas, carcinoma ductal *in situ* y cáncer invasivo fueron, respectivamente, 2,79, 1,99 y 1,36. Se encontró una relación positiva entre las tasas de realización de pruebas y las tasas de detección en punción-aspiración con aguja fina y lesiones benignas (R^2 : 0,53), punción-aspiración con aguja fina y carcinoma invasivo (R^2 : 0,28), biopsias cerradas y lesiones benignas (R^2 : 0,64), biopsias cerradas y carcinoma ductal *in situ* (R^2 : 0,61), y biopsias cerradas y carcinoma invasivo (R^2 : 0,48).

Conclusiones: Se observaron variaciones en la realización de pruebas invasivas entre las unidades de detección precoz de cáncer de mama de mayor magnitud que las de detección de lesiones. Las unidades con más pruebas de punción-aspiración con aguja fina tienen mayores tasas de detección de lesiones benignas, y las que realizan más biopsias cerradas detectan más lesiones benignas y cáncer.

© 2016 SESPAS. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la CC BY-NC-ND licencia (<http://creativecommons.org/licencias/by-nc-nd/4.0/>).

Variations in the diagnostic confirmation process between breast cancer mass screening units

A B S T R A C T

Keywords:

Breast cancer

Early detection of cancer

Diagnostic techniques and procedures

Clinical practice variation

Small area variation analysis

Assessment of health services

Objective: To analyse variations in the diagnostic confirmation process between screening units, variations in the outcome of each episode and the relationship between the use of the different diagnostic confirmation tests and the lesion detection rate.

Method: Observational study of variability of the standardised use of diagnostic and lesion detection tests in 34 breast cancer mass screening units participating in early-detection programmes in three Spanish regions from 2002–2011.

Results: The diagnostic test variation ratio in percentiles 25–75 ranged from 1.68 (further appointments) to 3.39 (fine-needle aspiration). The variation ratio in detection rates of benign lesions, ductal carcinoma *in situ* and invasive cancer were 2.79, 1.99 and 1.36, respectively. A positive relationship between rates of testing and detection rates was found with fine-needle aspiration-benign lesions (R^2 : 0.53), fine-needle aspiration-invasive carcinoma (R^2 : 0.28), core biopsy-benign lesions (R^2 : 0.64), core biopsy-ductal carcinoma *in situ* (R^2 : 0.61) and core biopsy-invasive carcinoma (R^2 : 0.48).

* Autor para correspondencia.

Correo electrónico: carmen.natal@sespa.es (C. Natal).

◇ GRUPO BELE (grupo de estudio de lesiones benignas): IMIM (Hospital del Mar), Barcelona: Andrea Burón, Xavier Castells, Josep María Corominas, Laia Domingo, Ana Rodríguez-Arana, Marta Román, Sonia Servitja, María Sala, Isabel Torá-Rocamora y Mar Vernet; Corporacio Sanitaria Parc Taulí, Sabadell (Barcelona): Marisa Bareand y Xavier Andreu; Instituto Catalán de Oncología, Barcelona: Lluïcia Benito y Carmen Vidal; Hospital de la Santa Creu i Sant Pau, Barcelona: Maria Jesus Quintana y Judit Sola-Roca; Dirección General de Salud Pública, Departamento de Salud, Gobierno de Cantabria, Santander: Mar Sánchez; Servicio de Salud del Principado de Asturias, Oviedo (Asturias): Carmen Natal; Universidad de Oviedo, Oviedo (Asturias); CIBER de Epidemiología y Salud Pública (CIBERESP), España: Adonina Tardón y Ana Fernández-Somoano; Institut d'Investigacions Sanitàries Pere Virgili, Reus (Tarragona): Jaume Galceran y Francina Saladie; Hospital Santa Caterina, Girona: Joana Ferrer; Plan Catalán del Cáncer, Barcelona Spanish Cancer Research Network (RTICC), Instituto de Salud Carlos III, Madrid: Josep Alfons Espinas.

Conclusions: Variation in the use of invasive tests between the breast cancer screening units participating in early-detection programmes was found to be significantly higher than variations in lesion detection. Units which conducted more fine-needle aspiration tests had higher benign lesion detection rates, while units that conducted more core biopsies detected more benign lesions and cancer.

© 2016 SESPAS. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La detección precoz del cáncer de mama en España está basada en programas de cribado poblacionales, con financiación pública, que siguen las recomendaciones de la Guía Europea de garantía de calidad en cribado y diagnóstico del cáncer de mama¹.

Las etapas del proceso de cribado poblacional son: 1) identificación e invitación personalizada de todas las mujeres de la población elegible; 2) realización de las pruebas de cribado; 3) realización de las pruebas de confirmación diagnóstica a las personas con alteraciones en las pruebas de cribado; y 4) tratamiento y seguimiento de las personas con lesiones confirmadas².

Aunque cada comunidad autónoma organiza las actividades del programa en su ámbito geográfico, hay una coincidencia en relación con la población diana, la convocatoria, la prueba de cribado y la periodicidad de esta. Los programas se dirigen a todas las mujeres entre 45/50 y 69 años de edad, con invitación personalizada, y la prueba de cribado es la mamografía con doble proyección (oblicua-media lateral y craneo-caudal) cada 2 años. Desde el año 2006, la cobertura poblacional en España es del 100%³.

Todos los programas clasifican las mamografías de cribado de acuerdo con el *Breast Imaging and Data System* del American College of Radiology (ACR BI-RADS®)⁴. Según esta clasificación, los resultados 1 (normal) y 2 (benigno) vuelven al cribado sin pruebas adicionales, y el resto (0, 3, 4 y 5) determinan una recitación para realizar la confirmación diagnóstica, con estudios adicionales de imagen o pruebas invasivas. En algunas circunstancias, después de un resultado sospechoso de lesión benigna (BI-RADS 3) en las pruebas de imagen, se opta por repetir la mamografía en un plazo inferior a los 2 años (estudios intermedios), o por realizar pruebas invasivas.

La homogeneidad descrita en las etapas 1 y 2 de la detección precoz no se mantiene en la confirmación diagnóstica, en la cual los programas tienen un comportamiento variable, con tasas de recitación (citaciones para evaluación) que se multiplican por 7 entre la comunidad autónoma con las tasas más altas y la que tiene las más bajas. La utilización de pruebas invasivas y de estudios intermedios también es variable, con tasas que se multiplican por 3,5 para la realización de pruebas invasivas y por 25 para la realización de estudios intermedios⁵. La variabilidad entre áreas sanitarias en las tasas de realización de tratamientos quirúrgicos, cirugía conservadora y no conservadora también es alta⁶.

Estas variaciones sugieren diferencias en los criterios de interpretación de la mamografía y de indicación de las pruebas diagnósticas. No se ha estudiado la relación que pueden tener estas diferencias con las tasas de detección de cáncer, ni la magnitud de las mismas entre unidades de cribado.

Los objetivos del presente estudio fueron analizar las variaciones en el proceso de confirmación diagnóstica entre unidades de cribado, evaluar si se produce una variabilidad en los resultados de cada episodio y analizar si hay alguna relación entre las tasas de las diferentes pruebas de confirmación diagnóstica y las tasas de detección de lesiones benignas y de cáncer de mama.

Métodos

Estudio observacional descriptivo de las variaciones en las tasas de realización de pruebas de confirmación diagnóstica de los programas poblacionales de detección precoz del cáncer de mama.

El ámbito de estudio fueron 34 unidades de cribado mamográficas. Estas unidades representan la totalidad de los programas de tres comunidades autónomas (Asturias, Cantabria y Cataluña) y el 25% de las unidades mamográficas fijas de España⁷. La unidad de cribado, que cuenta con recursos específicos, es el dispositivo sanitario que realiza las exploraciones propias del programa de detección precoz de cáncer de mama a una población diana definida.

El periodo de estudio fueron los años 2002 a 2011, ambos incluidos.

En cada episodio de cribado se recopila sistemáticamente información relacionada con la interpretación mamográfica, con la recitación para una evaluación adicional, cuya finalidad es descartar o confirmar la malignidad, y con las evaluaciones adicionales llevadas a cabo en cada caso. Con esta información se construyó una base de datos única y anonimizada, con datos individualizados para cada mujer que asistió al menos a una visita de cribado. El diagnóstico definitivo de cáncer de mama siempre se confirmó histopatológicamente utilizando los códigos de la Clasificación Internacional de Enfermedades para Oncología 10ª ed. (CIE-10). Los casos de cáncer de mama fueron clasificados como invasivos o carcinomas ductales *in situ* (CDIS).

Las medidas de resultado utilizadas para la realización del análisis de variabilidad en el proceso diagnóstico fueron las tasas estandarizadas por edad, de recitación, de realización de pruebas invasivas y de realización de cada tipo de prueba invasiva: punción-aspiración con aguja fina (PAAF), biopsia cerrada (BC) y biopsia quirúrgica (BQ).

Se realizó un análisis para los subgrupos BI-RADS 0, 3, 4 y 5 de las tasas de recitación y de realización de pruebas invasivas. Se excluyeron los BI-RADS 1 y 2 porque los estudios adicionales en estos tienen un carácter muy excepcional. No se realizó el análisis para los subgrupos BI-RADS en cada tipo de prueba invasiva, pues los grupos resultaron demasiado pequeños al desagregar. También se utilizaron las tasas estandarizadas por edad, de detección de lesiones benignas, detección de CDIS y detección de cáncer invasivo.

La tasa de recitación se definió como el número de citas para exploraciones adicionales después de una prueba de cribado; la tasa de realización de pruebas invasivas, como el número de exploraciones adicionales invasivas; las tasas de realización de PAAF, BC y BQ, como el número de cada una de las pruebas realizadas; la tasa de detección de lesiones benignas, CDIS y cáncer invasivo, como el número de diagnósticos de lesión benigna, CDIS o cáncer invasivo; todas ellas por cada mil episodios de cribado.

La estandarización por edad se realizó mediante el método directo utilizando como población de referencia la suma de las poblaciones de todas las unidades de cribado. Los estadísticos utilizados en el análisis de variabilidad fueron las razones de variación (RV) 5-95 y 25-75, y el componente sistemático de la variación (CSV)⁸.

Tabla 1

Descripción de la población de estudio: mujeres participantes, pruebas realizadas y resultados en las unidades de cribado estudiadas. Asturias, Cantabria y Cataluña. Años 2002-2011

Unidades de cribado	34	
Mujeres exploradas	523.480	
Episodios	1.521.720	
Recitaciones (por mil mujeres exploradas)	70.383	(46,3)
Pruebas invasivas (por mil mujeres exploradas)	16.145	(10,6)
Punción-aspiración con aguja fina (%)	6.882	(42,6)
Biopsia cerrada (%)	7.289	(45,1)
Biopsia abierta (%)	1.974	(12,2)
Episodios con lesiones benignas (tasa de detección ‰)	8.240	(15,7)
Episodios con cáncer de mama (tasa de detección ‰)	6.174	(11,8)
Carcinoma <i>in situ</i> (tasa de detección ‰)	1.050	(2,0)
Cáncer invasivo (tasa de detección ‰)	4.930	(9,4)

Para analizar el efecto del programa en la variabilidad de las tasas de realización de las diferentes pruebas y en las tasas de detección de lesiones se realizó un análisis de la varianza de un factor (ANOVA), considerando el programa la variable independiente.

Para la representación gráfica de la variación de las tasas analizadas se realizaron gráficos de puntos en escala logarítmica de media 0. Cada punto corresponde a una unidad de cribado. En el eje de abscisas se representó cada una de las pruebas analizadas, y en el eje de ordenadas se representó la desviación del logaritmo de la tasa estandarizada con respecto a la tasa media. Esta representación permite comparar en un mismo gráfico tasas con magnitudes muy diferentes.

Para medir la asociación entre las tasas de realización de pruebas invasivas y las tasas de detección se utilizó el coeficiente de regresión. Para la representación gráfica se elaboraron gráficos de burbujas que permiten incluir, además de la asociación entre las tasas analizadas, el volumen de actividad de cada unidad de cribado.

Para realizar los análisis se utilizó Stata 12.1® (Stata Corporation, College Station, Texas).

El estudio fue aprobado por todas las instituciones participantes y por el Comité de Ética de la Investigación Clínica de Salut MAR, del Hospital del Mar, sede del proyecto.

Resultados

Entre los años 2002 y 2011, en las 34 unidades de cribado estudiadas han participado 523.480 mujeres, con una media de

tres episodios de cribado. Fueron recitadas 46,3 mujeres de cada mil participantes, y a 10,6 mujeres de cada mil participantes se les realizó una prueba invasiva. El 43% de las pruebas invasivas fueron PAAF, el 45% BC y el 12% BQ. A 11,8 mujeres de cada mil participantes se les diagnosticó un CI, a 2 de cada mil un CDIS, y a 15,7 de cada mil una lesión benigna. Estos datos se presentan en la [tabla 1](#).

La [tabla 2](#) muestra los estadísticos de variabilidad para la tasa de recitaciones, la tasa de realización de pruebas invasivas y las tasas específicas de realización de PAAF, BC y BQ. Se encontró una RV_{25-75} de 1,68 para la tasa de recitación y de 1,81 para la realización de pruebas invasivas. En cuanto al tipo de prueba invasiva realizada, las razones de variación de las unidades de cribado comprendidas entre los percentiles 25 y 75, en las tasas de realización de pruebas invasivas, oscilaron entre 1,99 para la BC y 3,39 para la PAAF. Asimismo, la variación sistemática, no esperada por el azar, medida con el CSV en la tasa de recitación, fue menor que en la tasa de realización de pruebas invasivas (13% y 23%, respectivamente). En el análisis por subgrupos de pruebas invasivas, la menor variación sistemática se encontró en la tasa de realización de BC, con un CSV del 35%; para la tasa de realización de PAAF fue del 52%, y para la tasa de realización de BQ fue del 64%.

En la [figura 1](#) se observa la distribución de la variación de las tasas analizadas, con una menor dispersión para la tasa de realización de pruebas invasivas y la tasa de recitaciones que para las tasas de realización de cada una de las pruebas diagnósticas específicas.

En el análisis por subgrupos BI-RADS, la tasa de recitaciones presentó una RV_{25-75} de 1,63 para BI-RADS 0, de 5,94 para BI-RADS 3, de 1,08 para BI-RADS 4 y de 2,40 para BI-RADS 5. En el análisis por subgrupos BI-RADS de la tasa de realización de pruebas invasivas, las RV_{25-75} fueron de 2,10 para BI-RADS 0, 2,20 para BI-RADS 3, 1,50 para BI-RADS 4 y 2,38 para BI-RADS 5 (resultados no mostrados en las tablas).

Las variaciones en las tasas de detección (lesiones benignas, CDIS y cáncer invasivo) se presentan en la [tabla 3](#). Las mayores diferencias se encontraron en las tasas de detección de lesiones benignas, con una RV_{25-75} de 2,79, mientras que las RV_{25-75} de las tasas de detección, de carcinoma invasivo y de CDIS fueron <2. En la [figura 2](#) se representa esta variación; las tasas de detección de carcinoma invasivo de todas las unidades aparecen agrupadas en torno a los valores centrales y no se identificaron valores extremos, mientras que sí aparecen en las otras tasas analizadas.

Tabla 2

Variabilidad entre unidades de cribado poblacional de cáncer de mama en las tasas de realización de pruebas de confirmación diagnóstica ($\times 1.000$) basada en la estandarización directa. Años 2002-2011

	Recitaciones	Pruebas invasivas	Punción-aspiración con aguja fina	Biopsia cerrada	Biopsia quirúrgica
Episodios	70.383	14.215	6.882	7.289	1.974
Tasa cruda	46,25	9,34	4,52	4,79	1,30
Tasa estandarizada	42,52	8,22	3,39	4,79	1,29
Tasa estandarizada mínima	13,30	2,89	0,21	3,36	0,25
Tasa estandarizada máxima	74,15	20,20	10,96	1,21	4,04
Tasa estandarizada P ₅	18,01	4,08	1,07	11,95	0,53
Tasa estandarizada P ₂₅	32,40	2,27	2,11	1,40	0,83
Tasa estandarizada P ₅₀	39,14	6,49	3,00	2,09	1,29
Tasa estandarizada P ₇₅	54,59	9,55	7,17	3,39	2,19
Tasa estandarizada P ₉₅	69,29	17,19	10,90	4,15	3,67
RV ₂₅₋₇₅	1,68	1,81	3,39	1,99	2,6
RV ₅₋₉₅	3,85	4,21	10,19	7,40	6,94
CSV	0,13	0,23	0,52	0,35	0,64
ANOVA (variable independiente: programa de cribado)		0,88		0,91	0,57
		0,77-0,99		0,82-1,00	0,39-0,73
R ²	0,44	<0,0001	0,81	<0,0001	0,0012
IC95%	0,27-0,61		0,70-0,95		
p	0,02		<0,0001		

ANOVA: análisis de la varianza de un factor; CSV: coeficiente sistemático de la variación; IC95%: intervalo de confianza del 95%; P₅, P₂₅, P₅₀, P₇₅ y P₉₅: percentiles 5, 25, 50, 75 y 95; RV_{5-95} : razón de variación entre la tasa estandarizada de la unidad del percentil 95 y la de la unidad del percentil 5; RV_{25-75} : razón de variación entre la tasa estandarizada de la unidad del percentil 75 y la de la unidad del percentil 25.

Tabla 3
Variabilidad entre unidades de cribado poblacional de cáncer de mama en las tasas de resultados del episodio de cribado ($\times 1.000$) basada en la estandarización directa. Años 2002-2011

	Lesiones benignas	Carcinoma ductal <i>in situ</i>	Cáncer invasivo
Episodios	8.240	1.050	4.930
Tasa cruda	5,42	0,69	3,24
Tasa estandarizada	4,40	0,55	3,18
Tasa estandarizada mínima	0,54	0,07	2,03
Tasa estandarizada máxima	15,50	1,50	4,62
Tasa estandarizada P ₅	0,97	0,09	2,22
Tasa estandarizada P ₂₅	1,97	0,35	2,65
Tasa estandarizada P ₅₀	2,82	0,49	3,32
Tasa estandarizada P ₇₅	5,51	0,70	3,62
Tasa estandarizada P ₉₅	12,98	1,20	4,09
RV ₂₅₋₇₅	2,79	1,99	1,36
RV ₅₋₉₅	13,41	13,73	1,84
CSV	0,56	0,19	0,03
ANOVA			0,51
(variable independiente: programa de cribado)			0,33-0,67
R ²	0,91	0,85	0,0052
p	0,82-1,00	0,73-0,97	
IC95%	<0,0001	<0,0001	

ANOVA: análisis de la varianza de un factor; CSV: coeficiente sistemático de la variación; IC95%: intervalo de confianza del 95%; P₅, P₂₅, P₅₀, P₇₅ y P₉₅: percentiles 5, 25, 50, 75 y 95; RV₅₋₉₅: razón de variación entre la tasa estandarizada de la unidad del percentil 95 y la de la unidad del percentil 5; RV₂₅₋₇₅: razón de variación entre la tasa estandarizada de la unidad del percentil 75 y la de la unidad del percentil 25.

En el ANOVA realizado con el fin de controlar el efecto del programa al que están asignadas las unidades de cribado, los coeficientes R² oscilaron entre 0,91 para la tasa de realización de BC y 0,44 para la tasa de recitación; todos resultaron estadísticamente significativos (tablas 2 y 3).

Finalmente, en la figura 3 se representa la relación entre las diferentes pruebas invasivas realizadas y los resultados del cribado, según el número de episodios en cada unidad. En conjunto, las unidades con más actividad tienen tasas de realización de pruebas invasivas menores que las unidades más pequeñas. Se encontraron coeficientes de regresión estadísticamente significativos entre la tasa de realización de PAAF y la de detección de lesiones benignas y de carcinoma invasivo, aunque en esta última fue de 0,28. También resultaron estadísticamente significativas las regresiones entre las tasas de realización de BC y las tasas de detección de los tres tipos de lesiones estudiadas. Por último, la tasa de realización de BQ no guarda relación con los resultados del proceso de cribado.

Discusión

Se ha encontrado una variación sistemática en el proceso de confirmación diagnóstica de cáncer de mama en mujeres procedentes de programas poblacionales de cribado que puede clasificarse entre alta y muy alta, considerando que CSV superiores al 5% pueden considerarse una variación alta, y superiores al 10% muy alta⁹. Esta variación ocurre en la tasa de recitaciones, en la de realización de pruebas invasivas y en las tasas específicas de realización de PAAF, BC y BQ. La menor variabilidad se produce en la tasa de realización de BC, y la más alta en la tasa de realización de BQ. En el análisis por subgrupos BI-RADS, la mayor variabilidad correspondió a BI-RADS 3, tanto para la tasa de recitaciones como para la tasa de realización de pruebas invasivas. No se encontró que el programa al que estuvieran asignadas las unidades de cribado influyera de manera significativa en la variabilidad de las tasas analizadas.

La menor variabilidad en relación a las pruebas invasivas utilizadas corresponde a la BC, resultado esperable si se considera que es la prueba aceptada como de referencia en el diagnóstico

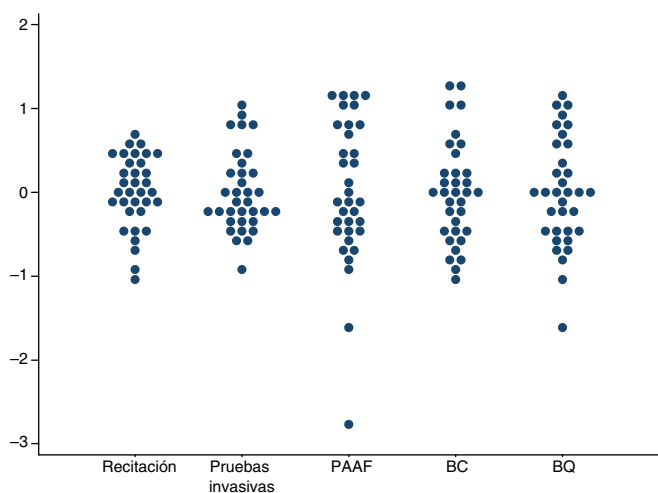


Figura 1. Variaciones en las tasas de realización de pruebas de confirmación diagnóstica por unidades de cribado de cáncer de mama. Años 2002-2011. Se representan las tasas en escala logarítmica centrada en el 0. Cada punto representa una unidad de cribado. BC: biopsia cerrada; BQ: biopsia quirúrgica; PAAF: punción-aspiración con aguja fina.

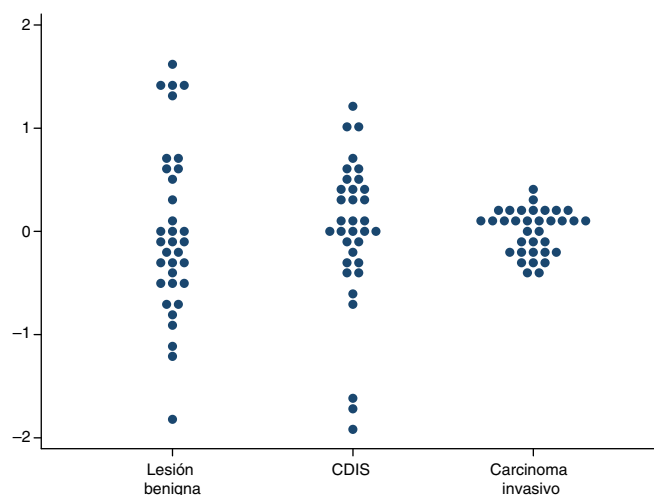


Figura 2. Variaciones en las tasas de detección por unidades de cribado de cáncer de mama. Años 2002-2011. Se representan las tasas en escala logarítmica centrada en el 0. Cada punto representa una unidad de cribado. CDIS: carcinoma ductal *in situ*.

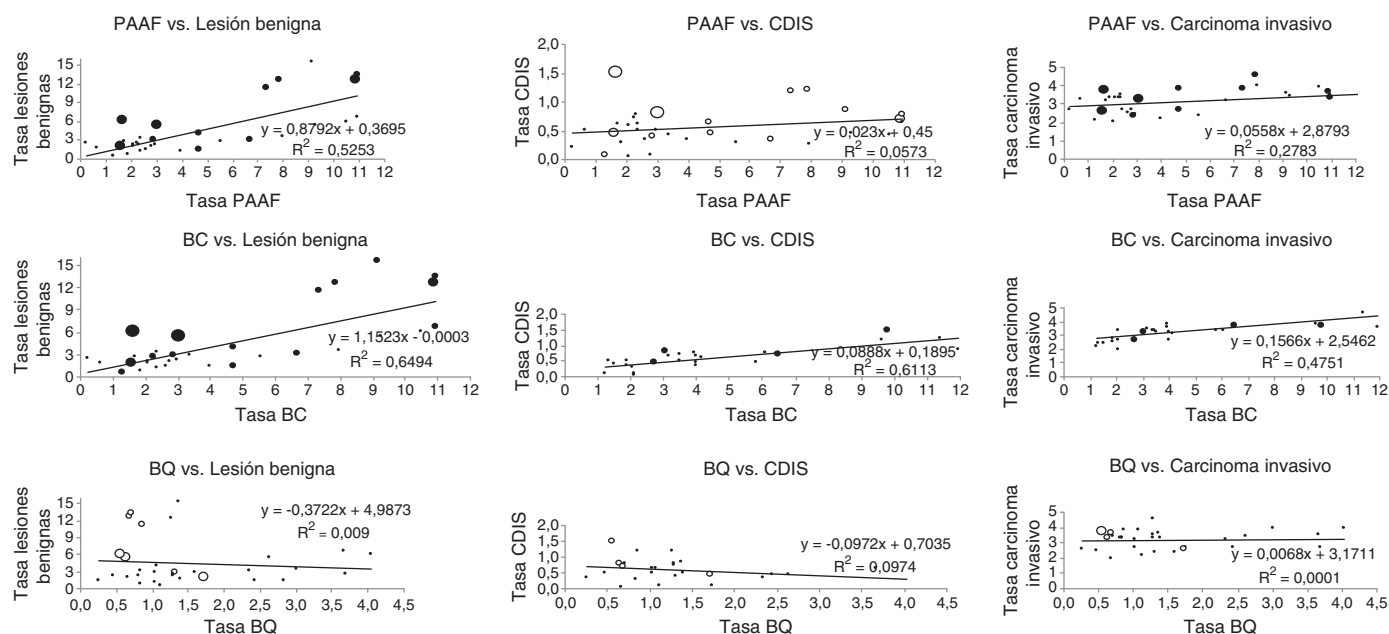


Figura 3. Relación entre las tasas (por 1000) de realización de pruebas invasivas y las tasas de detección por unidades de cribado de cáncer de mama. Años 2002-2011. Las burbujas negras representan coeficientes R^2 estadísticamente significativos, y las burbujas blancas representan coeficientes R^2 estadísticamente no significativos. BC: biopsia cerrada; BQ: biopsia quirúrgica; CDIS: carcinoma ductal *in situ*; PAAF: punción-aspiración con aguja fina.

de lesiones sospechosas de malignidad en el cribado mamográfico. La Guía Europea de Garantía de Calidad en cribado y diagnóstico del cáncer de mama¹ establece que una biopsia con aguja gruesa de calibre 1 realizada correctamente aporta mayor sensibilidad y especificidad que la PAAF cuando se trata de microcalcificaciones, asimetrías y distorsiones morfológicas. La BC también facilita el diagnóstico definitivo de las lesiones benignas. La PAAF podría ser el procedimiento referente en algunos centros para la toma de muestras en lesiones masivas y carcinomas obvios, pero sólo cuando se haya alcanzado un grado de excelencia satisfactorio. Finalmente, esta Guía recomienda que la biopsia abierta se reserve para cuando no sea posible realizar un diagnóstico preoperatorio. Se sabe que las tecnologías sujetas a mayor incertidumbre tienen mayor variabilidad¹⁰. Por otra parte, la variabilidad encontrada en la realización de BQ podría estar afectada porque esta prueba se realiza con menor frecuencia y su aceptabilidad es variable.

En relación con los resultados del episodio de cribado, la variabilidad entre unidades en las tasas de detección de cáncer, tanto CDIS como cáncer invasivo, fue baja, en especial para el tipo invasivo. Las mayores dificultades para la confirmación diagnóstica del CDIS en relación con el carcinoma invasivo, encontradas en otros estudios¹¹, podrían determinar más variabilidad en sus tasas de detección.

Se encontró una variabilidad moderada en las tasas de detección de lesiones benignas, que debe relacionarse con la mayor variabilidad en el manejo de las lesiones clasificadas como BI-RADS 3 (lesiones probablemente benignas) y con que se ha encontrado que las unidades que más PAAF realizan tienen una tasa de detección más alta de lesiones benignas, y las que realizan más BC tienen una mayor tasa de detección de los tres tipos de lesiones estudiadas.

En el estudio realizado se ha encontrado que hacer más recitaciones o pruebas invasivas no está asociado con la detección de cáncer de mama y sí con la detección de lesiones benignas. Aunque tanto el CDIS como las lesiones benignas tienen interés por su evolución a carcinoma invasivo^{12,13}, el objetivo de los programas de detección precoz de cáncer de mama es reducir la mortalidad

por cáncer de mama con los mínimos efectos adversos para las mujeres participantes, y la realización de estudios adicionales en los que finalmente no se diagnostica un cáncer de mama es el efecto adverso más frecuente de los mismos.

En todos los casos es destacable que los CSV son altos, lo que significa que las causas de esta variabilidad son sistemáticas y el efecto atribuible al azar es pequeño.

En este estudio debemos considerar ciertas limitaciones. En primer lugar, la variabilidad encontrada puede estar afectada, entre otros factores, por la variabilidad en la aplicación de la clasificación BI-RADS en la lectura de las mamografías, descrita en otros estudios¹⁴. Una segunda limitación es que en el análisis de los resultados no fue posible incluir las tasas de cáncer de intervalo (tumores no diagnosticados en el cribado que aparecen antes del cribado siguiente) de las distintas unidades analizadas, que representan una medida de los resultados falsos negativos de los programas; estas tasas podrían ser diferentes entre las unidades, aunque las informadas para los programas incluidos en este estudio están dentro de las recomendaciones^{15,16}. Tampoco se ha analizado la variabilidad para cada prueba desagregada por BI-RADS o por hallazgo radiológico debido a que los efectivos en cada grupo fueron muy pequeños. Los hallazgos radiológicos de la cohorte estudiada se han descrito en otros estudios¹⁷.

Por último, la representatividad de las unidades analizadas con relación a las disponibles en España es baja, y podría pensarse que hay programas con más variabilidad entre unidades que otros, aunque el bajo efecto del programa en la variabilidad de estas unidades encontrado en el estudio sugiere que la variabilidad se debe a otros factores diferentes a los programas.

Clásicamente se definen como factores vinculados a las variaciones en la práctica médica las deficiencias en la fiabilidad del análisis, los factores asociados a la demanda y los factores asociados a la oferta, considerando tanto la organización como el proveedor directo¹⁰; en este caso, es el profesional quien decide qué procedimiento diagnóstico utilizar.

Las posibles deficiencias en la fiabilidad del análisis se han descrito en las limitaciones del estudio. En cuanto a los factores asociados a la demanda, se asume que los programas poblacionales

eluden su influencia, ya que la oferta es universal y determinada por el sistema sanitario y no por la demanda individual.

Se dispone de pocos estudios que analicen el proceso diagnóstico en el contexto de los programas poblacionales de cribado¹⁸; la mayor parte analizan los determinantes sociales^{19–21} o las creencias de los profesionales^{22–25}, o la disponibilidad de recursos²⁶, en la participación y la cobertura de los programas. Este estudio permite tener la unidad de cribado como unidad de análisis. Cada programa cuenta con varias unidades de cribado, en distintos centros sanitarios, con diferentes culturas y modelos de gestión y planificación diferentes. Estas circunstancias podrían repercutir en los procedimientos diagnósticos incluidos en los programas poblacionales de cribado, y no necesariamente contemplados en su diseño y evaluación.

Los programas de cribado poblacional han realizado y realizan un gran esfuerzo por seguir las recomendaciones de las guías de práctica clínica en cuanto a la prueba de cribado, la población diana y la periodicidad de las pruebas, resultando una buena homogeneidad de criterios en todos ellos⁵. Este esfuerzo también es necesario en el proceso diagnóstico, en el cual los programas poblacionales de cribado y los hospitales implicados deben promover el uso de las tecnologías que resulten más eficientes, mediante la dotación de recursos, la formación de los profesionales, la evaluación del proceso asistencial y la difusión de las guías clínicas.

Editora responsable del artículo

M^a Felicitas Domínguez Berjón.

Declaración de transparencia

La autora principal (garante responsable del manuscrito) afirma que este manuscrito es un reporte honesto, preciso y transparente del estudio que se remite a GACETA SANITARIA, que no se han omitido aspectos importantes del estudio, y que las discrepancias del estudio según lo previsto (y, si son relevantes, registradas) se han explicado.

¿Qué se sabe sobre el tema?

Los estudios sobre cribado poblacional de cáncer de mama se han centrado en el análisis de los procesos de cribado y de evaluación de resultados, pero se dispone de poca información sobre el proceso de confirmación diagnóstica.

¿Qué añade el estudio realizado a la literatura?

Esta investigación estudia la variabilidad del proceso de confirmación diagnóstica en el cribado poblacional de cáncer de mama, y encuentra una variabilidad alta en la utilización de las pruebas diagnósticas. Para disminuir la variabilidad, los programas poblacionales de cribado de cáncer de mama y los hospitales implicados en el proceso diagnóstico deben facilitar la introducción temprana de las tecnologías más eficientes.

Financiación

Este trabajo ha sido financiado con fondos del Instituto de Salud Carlos III FEDER, (PI11/01296). Las fuentes de financiación no han participado en la realización del trabajo ni en la preparación del manuscrito.

Contribuciones de autoría

Todos/as los/as autores/as han participado en la idea, la concepción, el diseño y la elaboración de este artículo. Específicamente, C. Natal, A. Tardón y X. Castells concibieron el artículo; C. Natal, X. Castells e I. Torá-Rocamora recopilaron los datos; A. Fernández-Somoano realizó el análisis estadístico y la interpretación de los datos. Todos/as los/as autores/as han leído y aprobado la versión final.

Conflicto de intereses

Ninguno.

Bibliografía

- Perry N, Broeders M, de Wolf C, et al., editores. European Guidelines for quality assurance in breast cancer screening and diagnosis. European Commission. 4th ed. Luxembourg: Office for Official Publications of the European Communities; 2006.
- Anttila A, Lönnberg S, Ponti A, et al. Towards better implementation of cancer screening in Europe through improved monitoring and evaluation and greater engagement of cancer registries. *Eur J Cancer*. 2015;51:241–51.
- Castells X, Sala M, Asuncion N, et al., coordinadores. Descripción del cribado del cáncer en España. Proyecto DESCRIC. Madrid: Plan de Calidad para el Sistema Nacional de Salud. Ministerio de Sanidad y Consumo. Agència d'Avaluació de Tecnologia i Recerca Mèdiques de Catalunya; 2007. Informes de Evaluación de Tecnologías Sanitarias, AATRM núm. 2006/01.
- Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS® Mammography. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology; 2013.
- Asuncion N, Delfrade J, Salas D, et al. Programas de detección precoz de cáncer de mama en España: características y principales resultados. *Med Clin (Barc)*. 2013;141:13–23.
- Angulo-Pueyo E, Ridao-López M, Martínez-Lizaga N, et al. Variabilidad y coste de oportunidad de las alternativas quirúrgicas en cáncer de mama. *Gac Sanit*. 2014;28:209–14.
- Almazán R, Fernández AB, Zubizarreta R. Situación de los programas de cáncer de mama en España. 2008. XIII Reunión anual de los programas de detección precoz de cáncer de mama. Ámbito ibérico. Palma de Mallorca, 2009. [Consultado 25 Jul 15]. Disponible en: <http://www.cribadocancer.es/images/archivos/mama/situacion/Mama2008.pdf>
- Martínez-Lizaga N, Montes Y, Rodrigo I, et al. Metodología del atlas de variaciones en hospitalizaciones potencialmente evitables en el Sistema Nacional de Salud. *Atlas Var Pract Med Sist Nac Salud*. 2011;4:371–8.
- McPherson K. Systematic variation in surgical procedures and hospital admission rates: a methodological study: report II, April 1996, for the Department of Health. London: London School of Hygiene & Tropical Medicine, Dept. of Public Health & Policy, Health Promotion Sciences Unit; 1996.
- Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med*. 1982;16:811–24.
- Erbas B, Provenzano E, Armes J, et al. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat*. 2006;97:135–44.
- Duffy SW, Diben A, Michalopoulos D, et al. Screen detection of ductal carcinoma in situ and subsequent incidence of invasive interval breast cancers: a retrospective population-based study. *Lancet Oncol*. 2016;17:109–14.
- Castells X, Domingo L, Corominas JM, et al. Breast cancer risk after diagnosis by screening mammography of nonproliferative or proliferative benign breast disease: a study from a population-based screening program. *Breast Cancer Res Treat*. 2015;149:237–44.
- Boyer B, Canale S, Arfi-Rouche J, et al. Variability and errors when applying the BI-RADS mammography classification. *Eur J Radiol*. 2013;82:388–97.
- Prieto García MA, Delgado Sevillano R, Baldó Sierra C, et al. Clasificación y características del cáncer de intervalo en el programa de detección precoz de cáncer de mama en el Principado de Asturias. *Radiología*. 2013;55:408–15.
- Blanch J, Sala M, Ibáñez J, et al. Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. *PLoS One*. 2014;9:e110207.
- Castells X, Torá-Rocamora I, Posso M, et al. Risk of breast cancer in women with false-positive results according to mammographic features. *Radiology*. 2016;151174 [Epub ahead of print].
- Corallo AN, Croxford R, Goodman DC, et al. A systematic review of medical practice variation in OECD countries. *Health Policy*. 2014;114:5–14.
- McDonald JT, Sherman A. Determinants of mammography use in rural and urban regions of Canada. *Can J Rural Med*. 2010;15:52–60.
- Khan N, Kaestner R, Salmon JW, et al. Does supply influence mammography screening? *Am J Health Behav*. 2010;34:465–75.

21. Coughlin SS, Leadbetter S, Richards T, et al. Contextual analysis of breast and cervical cancer screening and factors associated with health care access among United States women, 2002. *Soc Sci Med.* 2008;66:260–75.
22. Symons AB, Mahoney MC, Englert J, et al. Variations in approaches to breast cancer screening among primary care physicians. *J Cancer Educ.* 2002;17:205–10.
23. Goel V, Iron K, Williams I. Enthusiasm or uncertainty: small area variations in the use of mammography services in Ontario, Canada. *J Epidemiol Community Health.* 1997;51:378–82.
24. Santora LM, Mahoney MC, Lawvere S, et al. Breast cancer screening beliefs by practice location. *BMC Public Health.* 2003;3:9.
25. Haggerty J, Tudiver F, Brown JB, et al. Patients' anxiety and expectations: how they influence family physicians' decisions to order cancer screening tests. *Can Fam Physician.* 2005;51:1658–9.
26. Autier P, Ait Ouakrim D. Determinants of the number of mammography units in 31 countries with significant mammography screening. *Br J Cancer.* 2008;99:1185–90.